Applications of QSAR to Drug Metabolizing Enzymes

Sean Ekins

ACT LLC, New York, NY
& School of Pharmacy, Department of Pharmaceutical Sciences, University of Maryland

…mathematical learning will be the distinguishing mark of a physician from a quack…

Richard Mead
A mechanical account of poisons in several essays
2nd Edition, London, 1708.

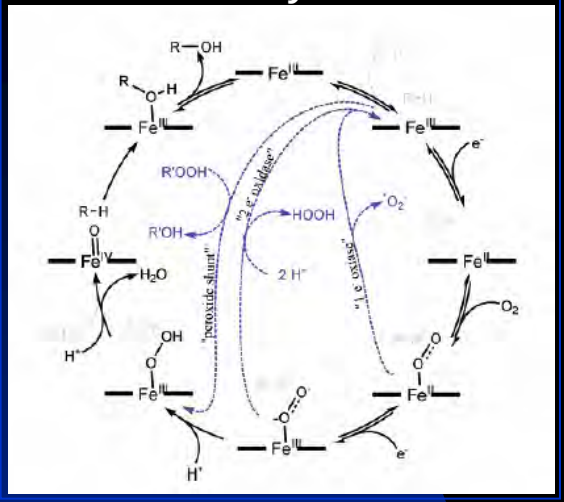
"….*drug discovery & development needs to be more like engineering"*
Janet Woodcock, FDA – PharmaDiscovery May 10 2006

Metabolism Then & Now

- **Hippuric acid formation from benzoic acid (Keller 1842)**
- Metabolism of 1000s of compounds assessed daily
  - Sensitivity of analytical tools increased
  - Many “minor” metabolites identified
- However data in public domain is sparse
- Focus on just a few enzymes – well characterized
- How can we improve throughput ?
- How can we use the metabolism data to predict toxicity?

Cytochrome P450

Reaction cycle



Which P450s

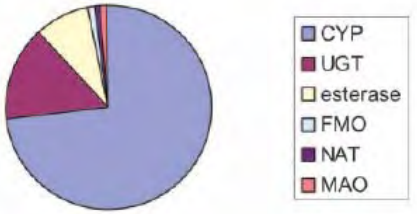
Regiospecificity | Lability | Affinity | Induction | Inhibition

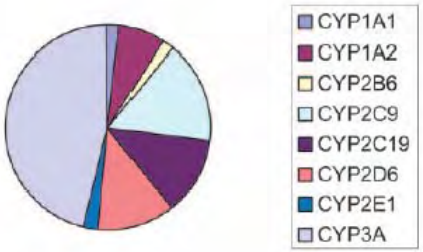
Site of metabolism

Rate of metabolism, V_{max}

K_m, K_d

Involvement in metabolism





After van de Waterbeemd & Gifford Nat Revs Drug Disc 2:192-204 (2003)
Williams et al DMD 32:1201-8 (2004), de Graaf et al., J Med Chem 48: 2725-2755 (2005)

Computational approaches

Ligand based

Quantitative structure activity relationship (QSAR), pharmacophore

Protein based

Homology models, docking, molecular dynamics simulations

Rule based

MetabolExpert (Darvas et al), **META** (Klopman et al), **Meteor** (LHASA)

Metabolism databases

Metabolite (MDL), **Metabolism** (Accelrys) **Assign occurrence frequencies to metabolites**

Combined/hybrid methods

MetaSite (Cruciani et al) Site of metabolism prediction for CYP2C9 and CYP3A4 etc

MetaDrug, Combining similarity to known ligands and regulatory and metabolic pathways, QSAR models etc.

Ekins et al., Expert Opin Drug Metab Toxicol 1: 303-323 (2005), de Graaf et al., J Med Chem 48: 2725-2755 (2005), Locuson and Wahlstrom DMD 33:873–878 (2005)

3D-QSAR

- A pharmacophore is the geometric arrangement of functional groups necessary for a biological response

- Assumes molecules bind and orient similarly in same active site & Pharmacophore represents common features of ligands

- Comparative molecular field analysis (CoMFA)

- Catalyst (Accelrys)

Pharmacophore Methods

Generate data > 16 molecules in vitro, Kd, Ki

- Activities should span 4 orders of magnitude
- Each magnitude should be represented by 3 compounds
- No redundant information
- No excluded volume problems

Generate 3D conformations of molecules

Align molecules

Select features contributing to activity

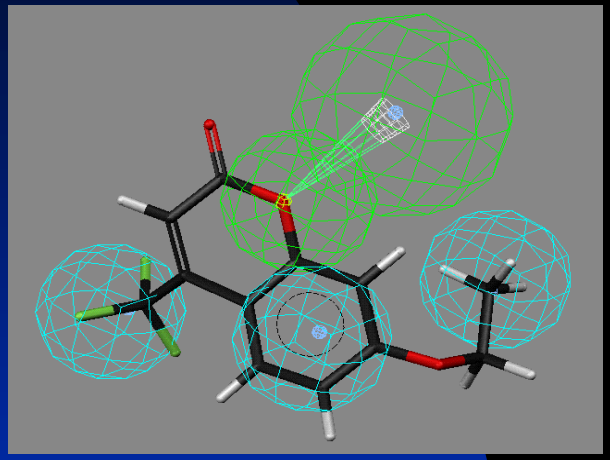
Regress hypothesis

Evaluate with new molecules with in vitro data

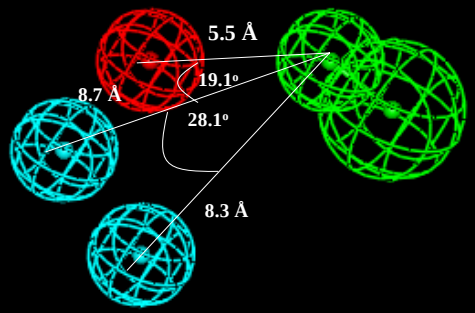
Result – 3D model that new molecules can be tested with

CYP Substrate Affinity Pharmacophores

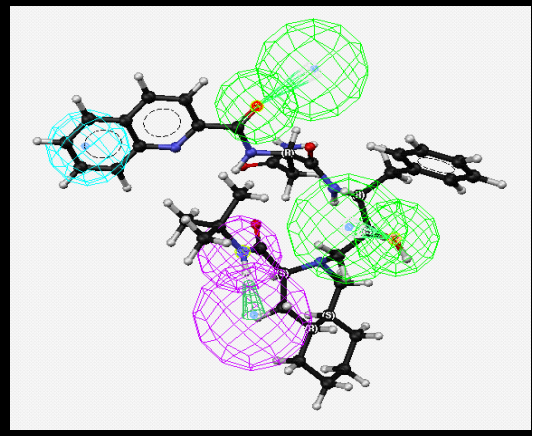
CYP2B6



CYP2D6



CYP3A4

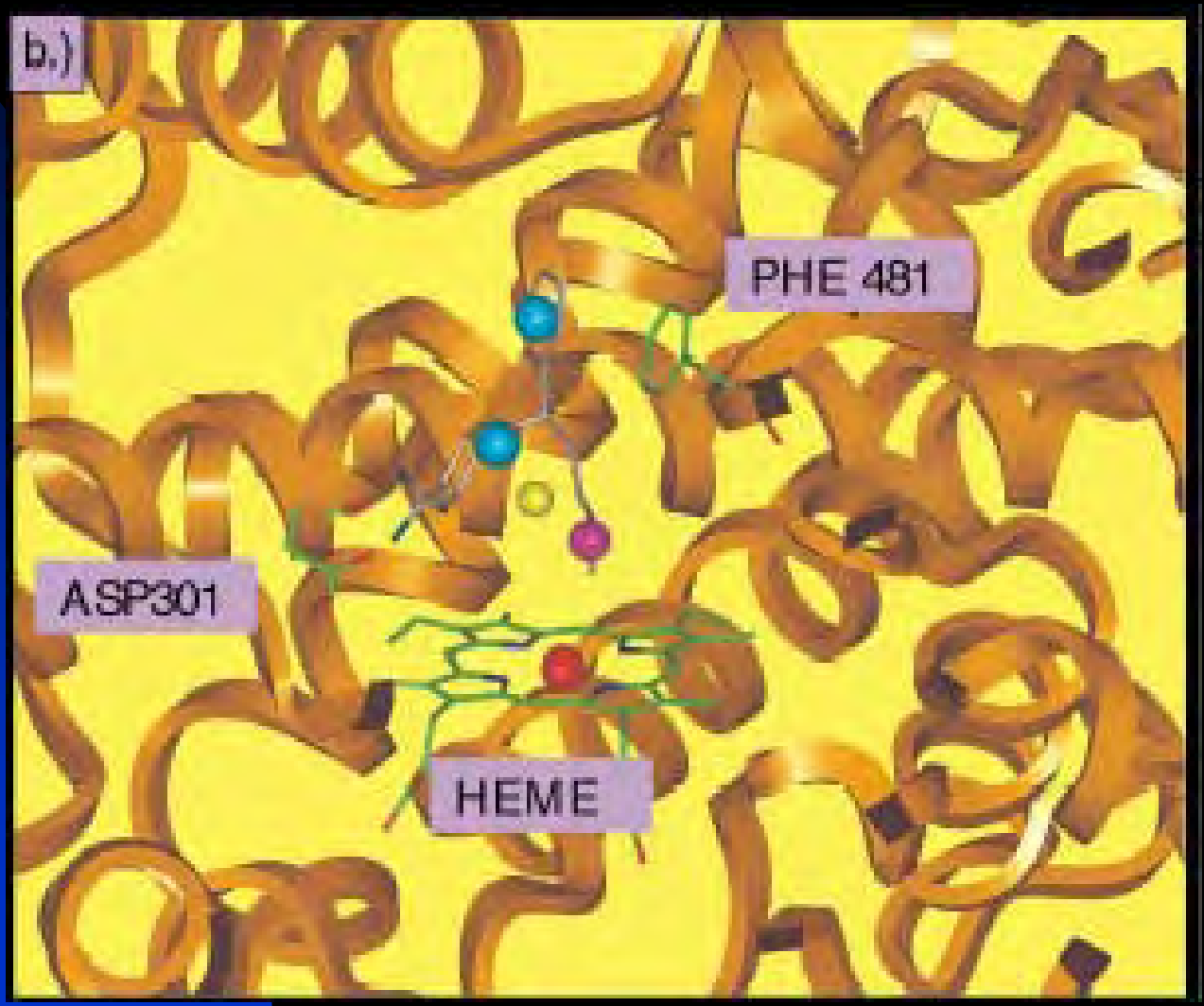


Ekins et al., JPET, **288**:21-29, (1999)
Ekins et al., JPET, **291**:424-433, (1999)
Ekins, S., de Groot, M. & Jones, J. P. DMD **29**, 936-944, (2001)
Wang & Halpert, DMD **30**: 86-95, (2002)
Snyder et al., QSAR, **21**: 357-368, (2002)

Integrated Pharmacophore and Homology Model

Inside CYP2D6: Homology model based on rabbit CYP2C3/5

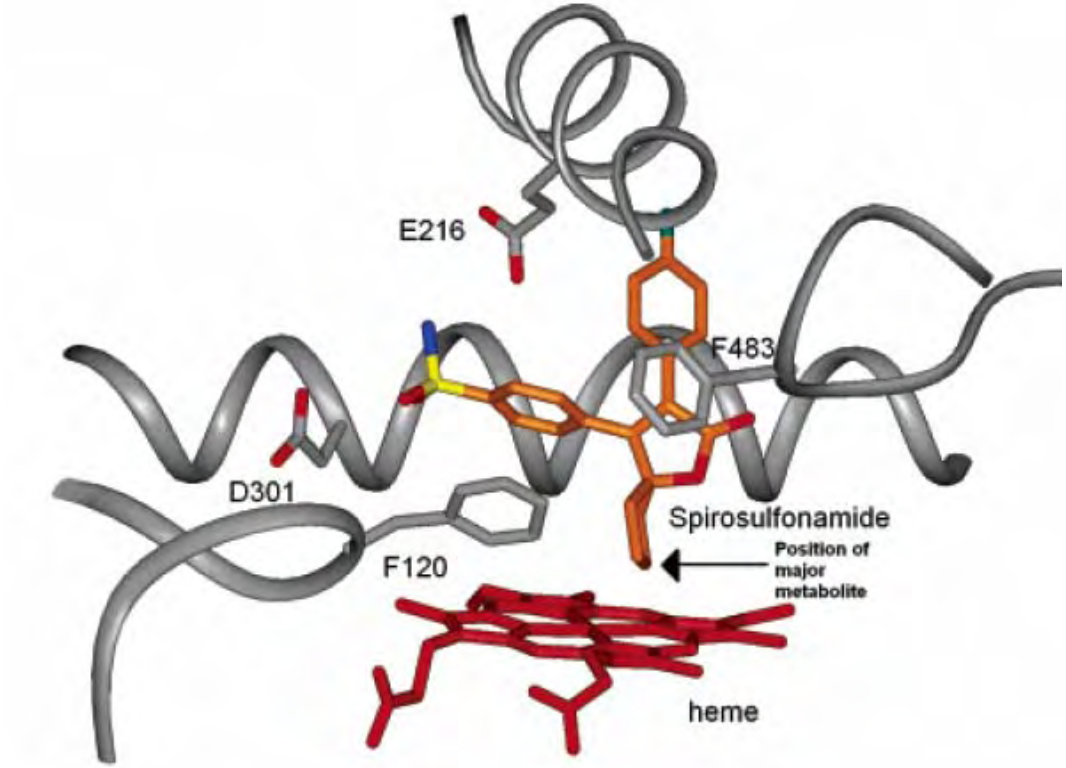
Fluoxetine -Showing position for N-demethylation

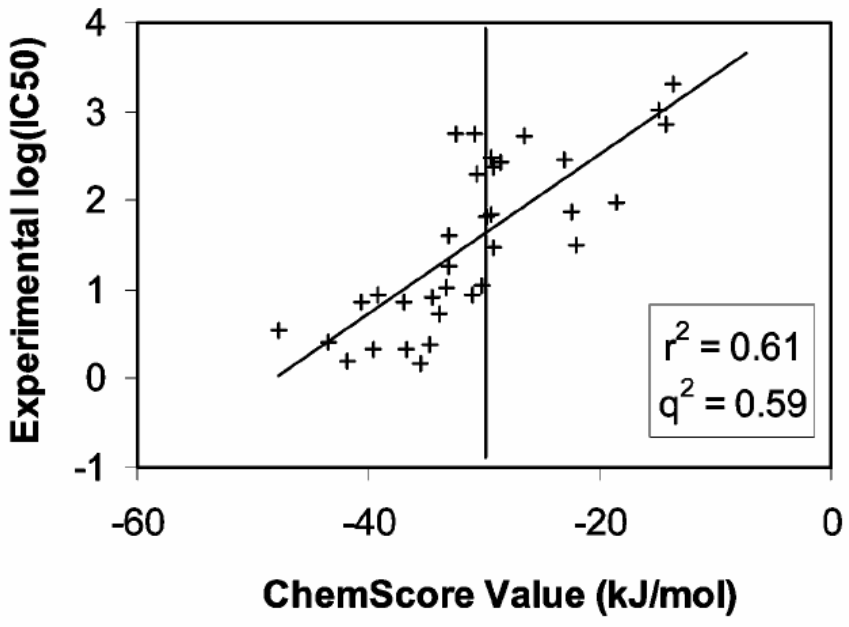


Snyder et al, QSAR 21: 357- 368 (2002)

CYP2D6 homology model Docking & Scoring

- **CYP2D6 homology model**
- **Docked & scored NCI compounds**
- **Generated experimental data for CYP2D6 inhibition (IC50)**
- **To date no information on substrates and affinity vs ChemScore**
- Kemp et al., *J. Med. Chem.* 2004, *47,* 5340-5346





CYP3A4 Summary

Dominant enzyme in drug metabolism -**an inducible enzyme** -catalytic activity highly variable - Expressed in Liver, Kidney and GI tract

Broad substrate specificity implies large active site

Metabolizes many classes of drugs / opportunities for DDI

Bulky hydrophobic groups present on substrates

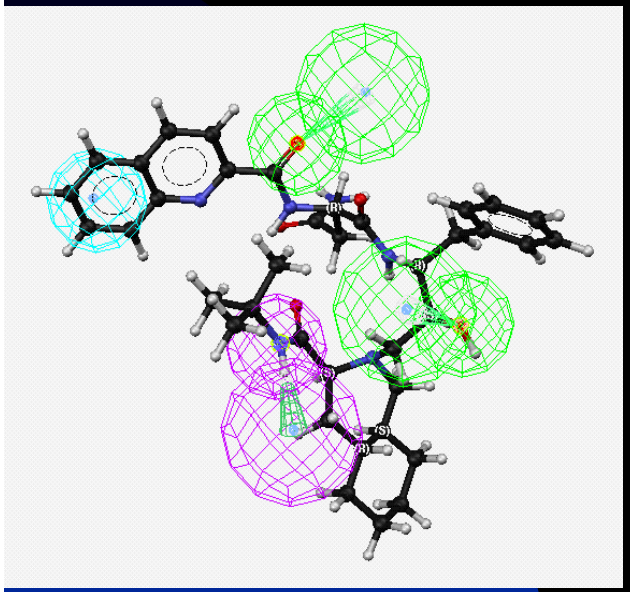
Some AA residues identified for inhibitor binding in active site

3D-QSAR models & many homology models

Szklarz and Halpert, J Comp Aided Mol Design 11; 265-272 (1997)

CYP3A4 Pharmacophores

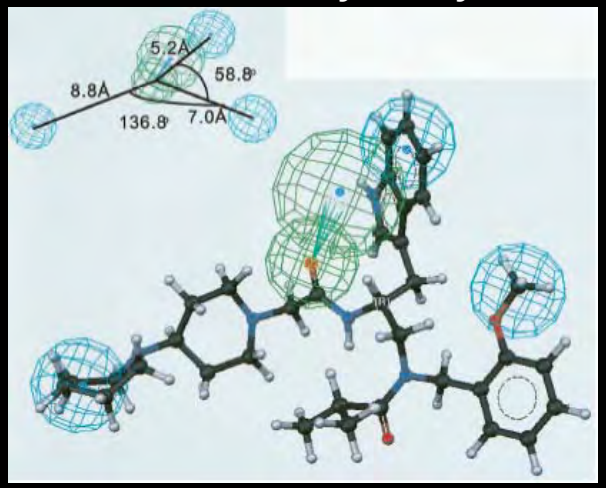
Substrate



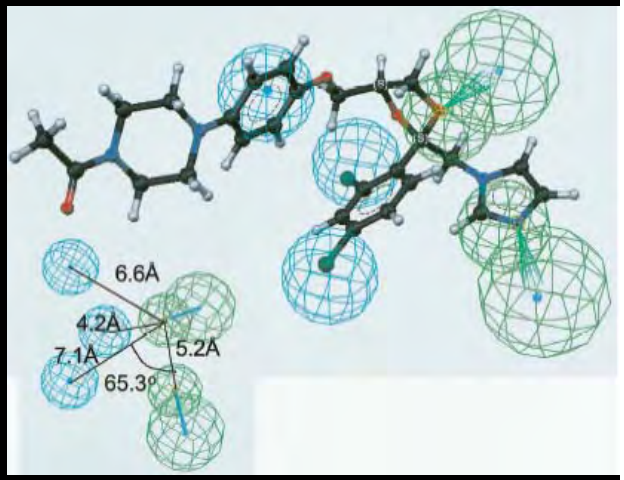
Ekins et al., JPET, **291**:424-433, (1999)

Inhibitor models

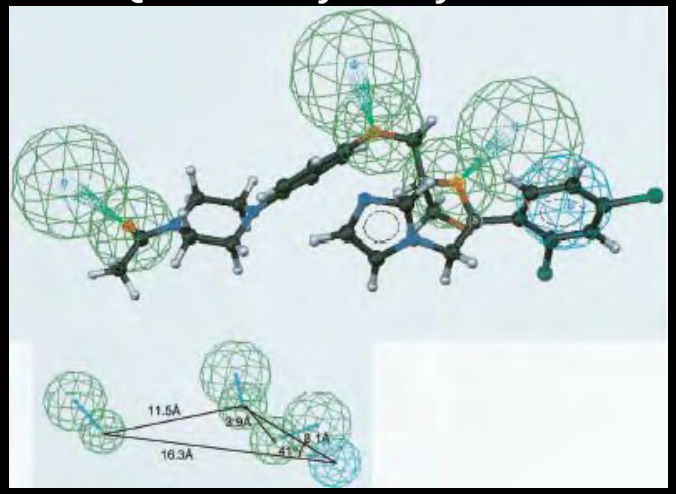
Midazolam 1'-hydroxylation



Cyclosporin A Metabolism

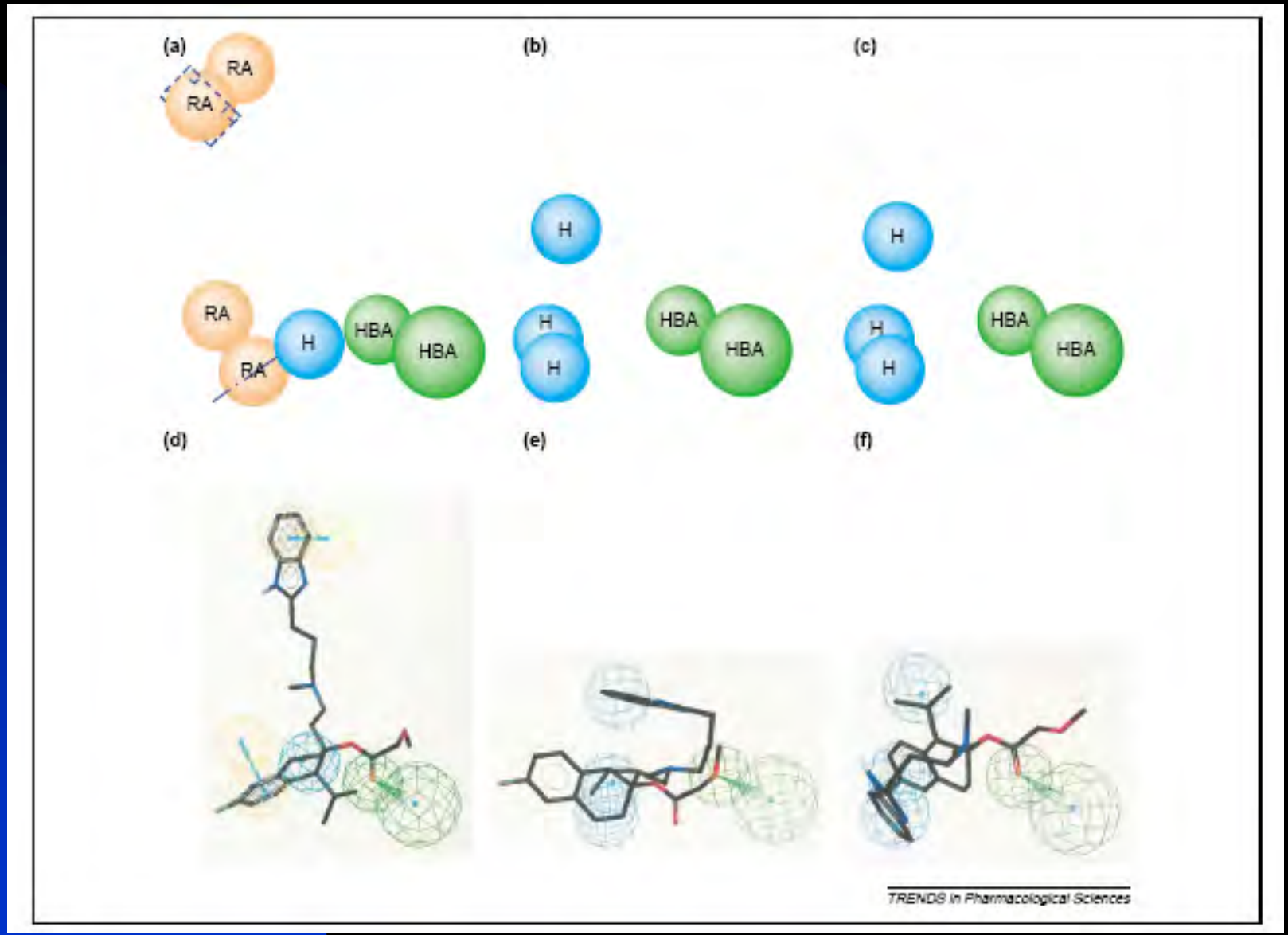


Quinine hydroxylation



Ekins et al., JPET, **290**:429-438, (1999)

CYP3A4, CYP3A5, CYP3A7 Inhibitor pharmacophores

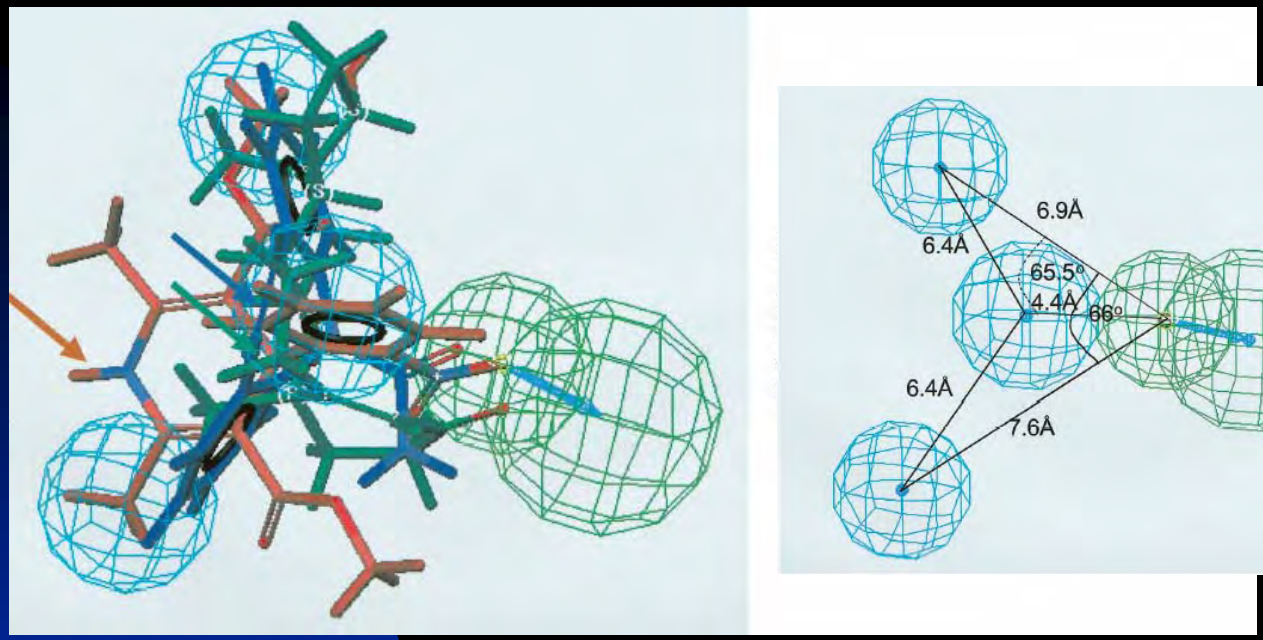


Ekins et al TIPS (2003) 24:161-166

CYP3A4 Autoactivators and Heteroactivators

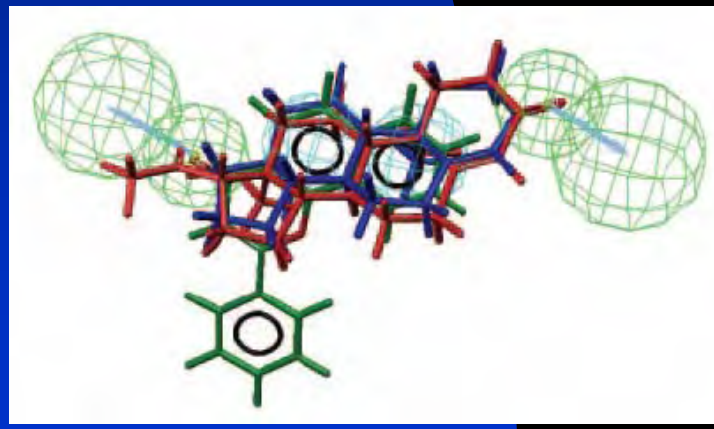
Autoactivators

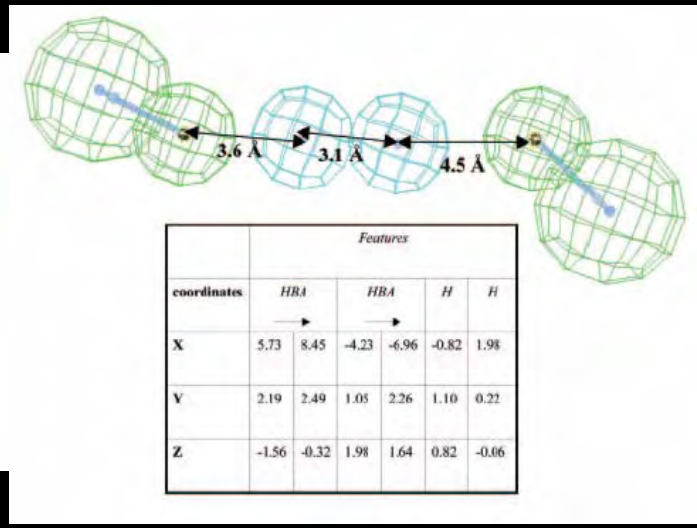
Ekins et al., JPET, **291**:424-433, (1999)



Testosterone
Nifedipine
carbamazepine

Heteroactivators



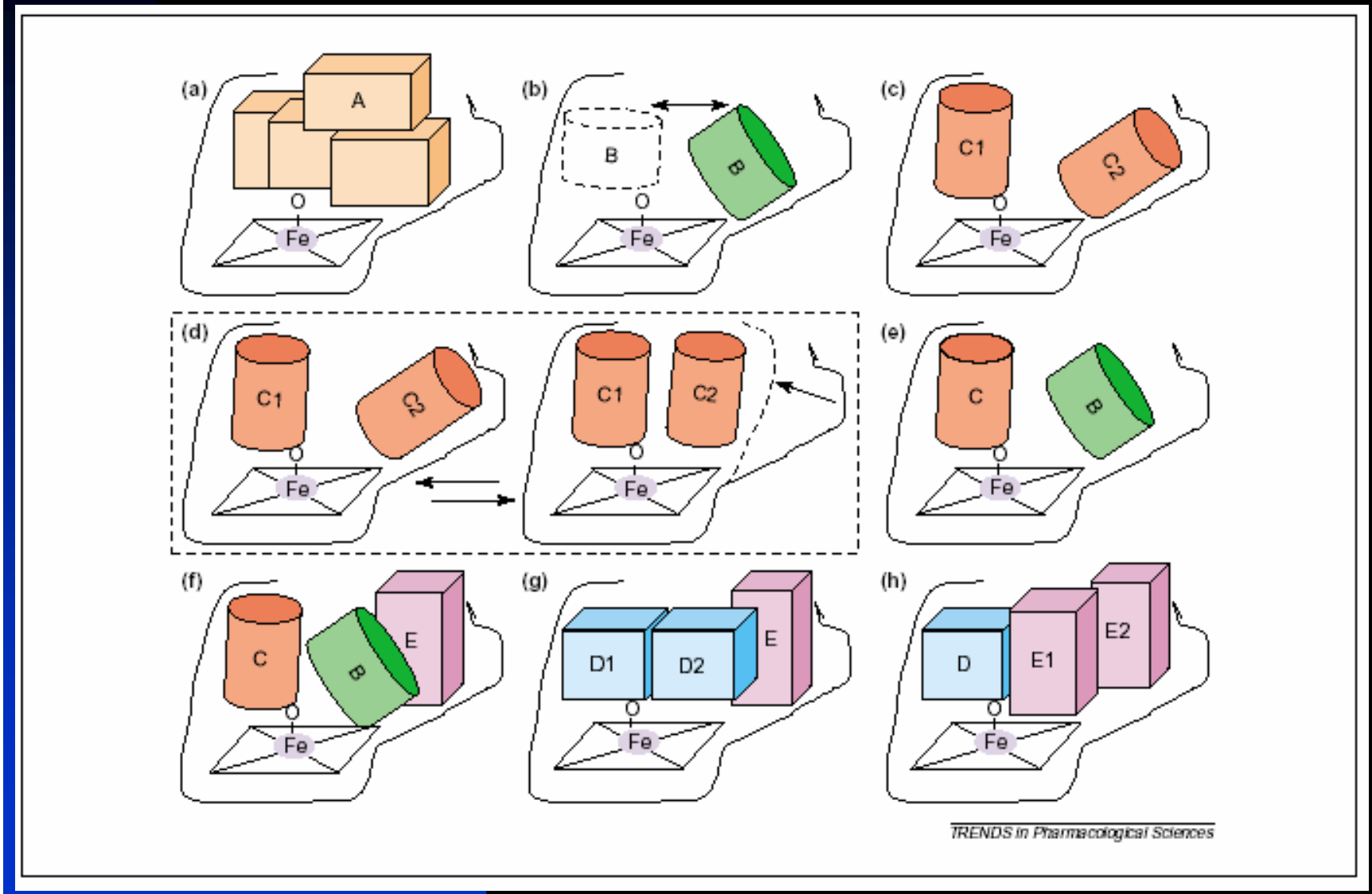


	Features					
coordinates	HBA →		HBA →		H	H
X	5.73	8.45	-4.23	-6.96	-0.82	1.98
Y	2.19	2.49	1.05	2.26	1.10	0.22
Z	-1.56	-0.32	1.98	1.64	0.82	-0.06

Testosterone
A-napthoflavone
Progesterone
Artemisinin
Quinidine
felbamate

Egnell et al., JPET 312:926–937, 2005

CYP3A4 binding site Hypotheses

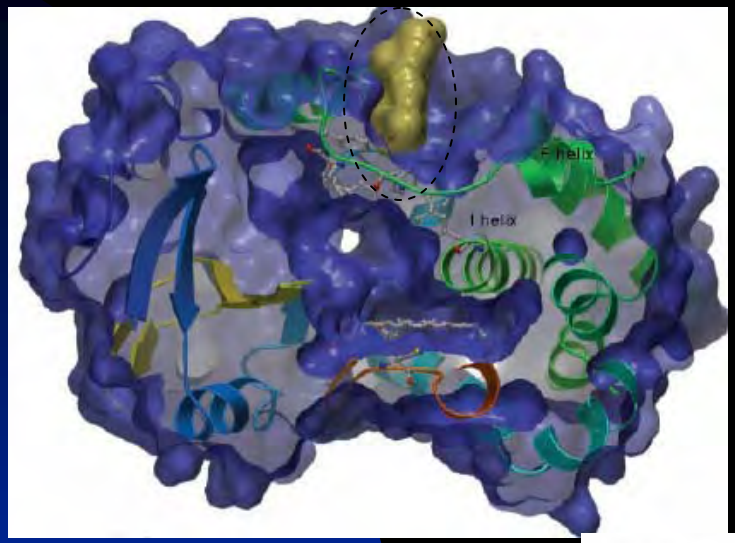


Ekins et al TIPS (2003) 24:161-166

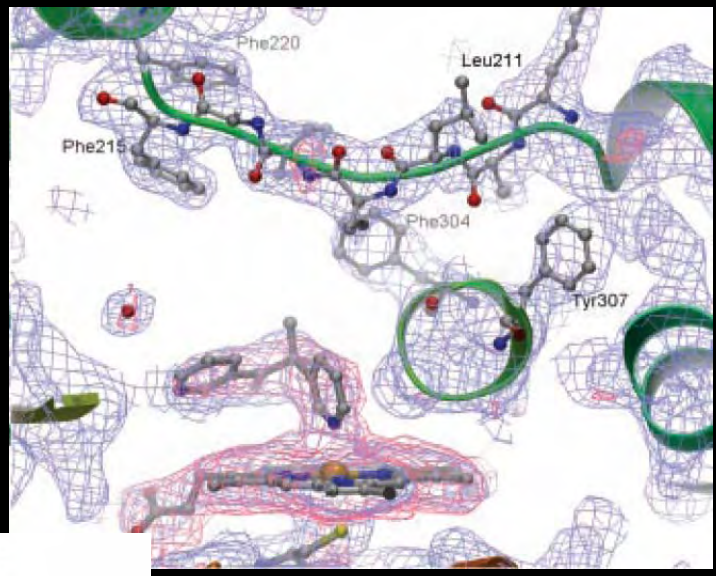
Human CYP3A4 X-ray structures

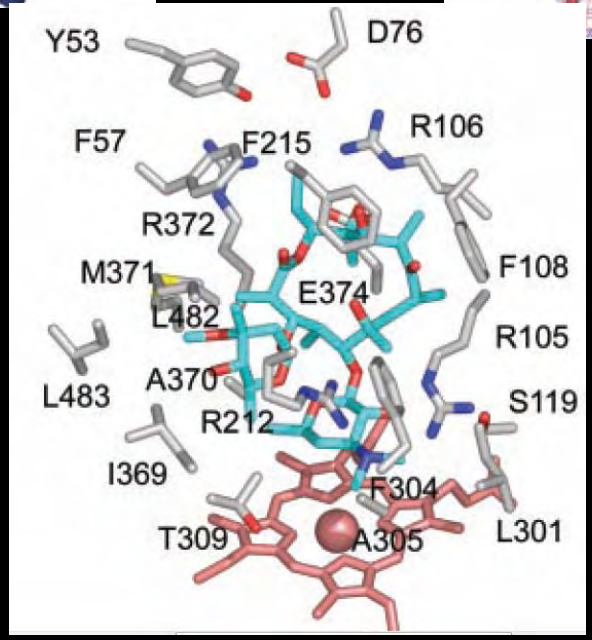
Williams et al., Science 305: 683-686 (2005)

Progesterone



Metyrapone





Erythromycin (docked)

Yano et al., J Biol Chem 279, 38091–38094, 2004

Reference Database of CYP substrates

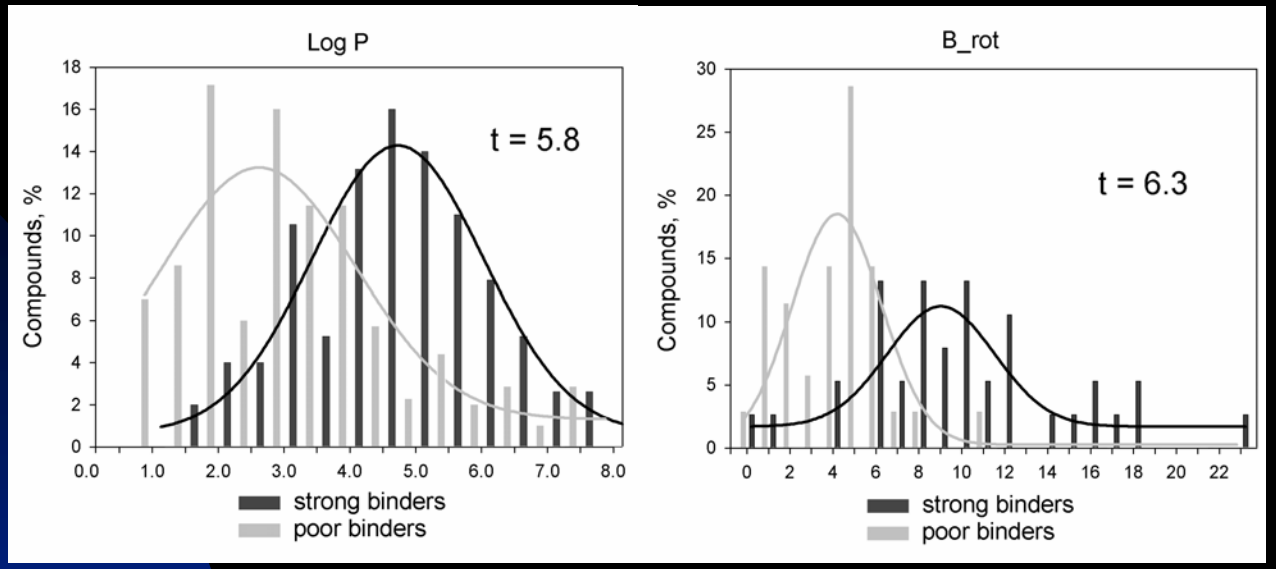
Collected K_m data for CYPs from the literature
Split data into groups

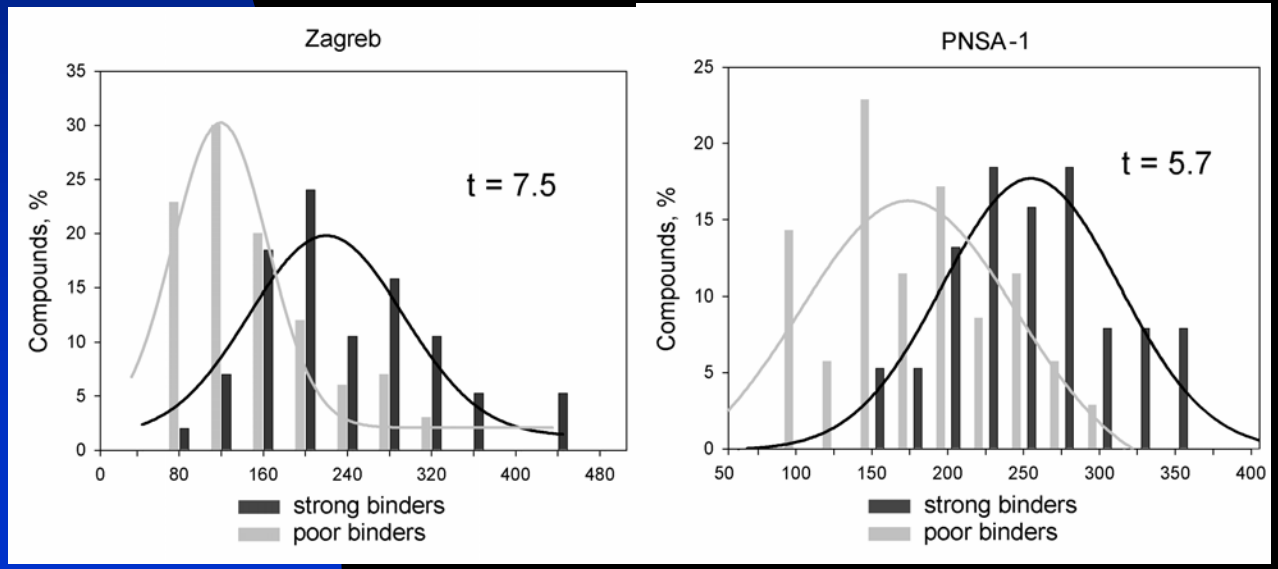
Enzyme	no. of compounds	no. of compounds		
		K_m<10	K_m=10-100	K_m>100
CYP1A1	12	7	4	1
CYP1A2	43	17	16	10
CYP2A6	15	1	3	11
CYP2B6	51	15	19	17
CYP2C8	13	6	5	2
CYP2C9	41	12	21	8
CYP2C19	48	18	21	9
CYP2D6	75	45	23	7
CYP2E1	19	2	8	9
CYP3A4	126	38	56	32
CYP3A5	12	5	6	1
CYP19	18	18	0	0
Total	491	180	208	103

Descriptor	Definition
LogP	log of 1-octanol/water partition coefficient
B_rot	Number of rotatable bonds
HBA	Number of H-bond acceptors
HBD	Number of H-bond donors
PNSA-1	Partial negative surface area
Zagreb	Sum of the squares of vertex valencies

Balakin et al DMD 32: 1183-1189 (2004)

Differences for low and high Km CYP3A4 binders (n=126)





Balakin et al DMD 32: 1183-1189 (2004)

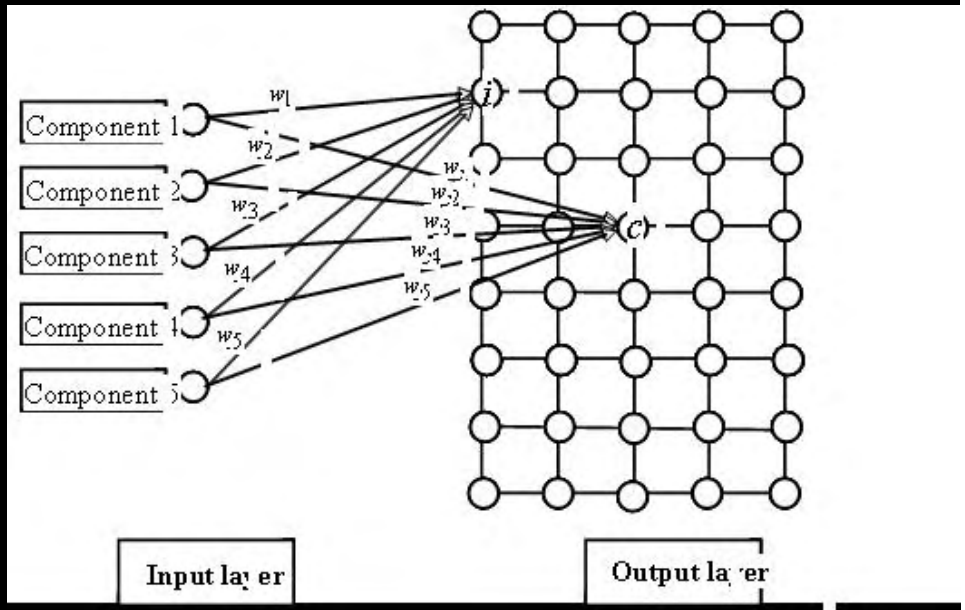
Self-Organizing Map (SOM)

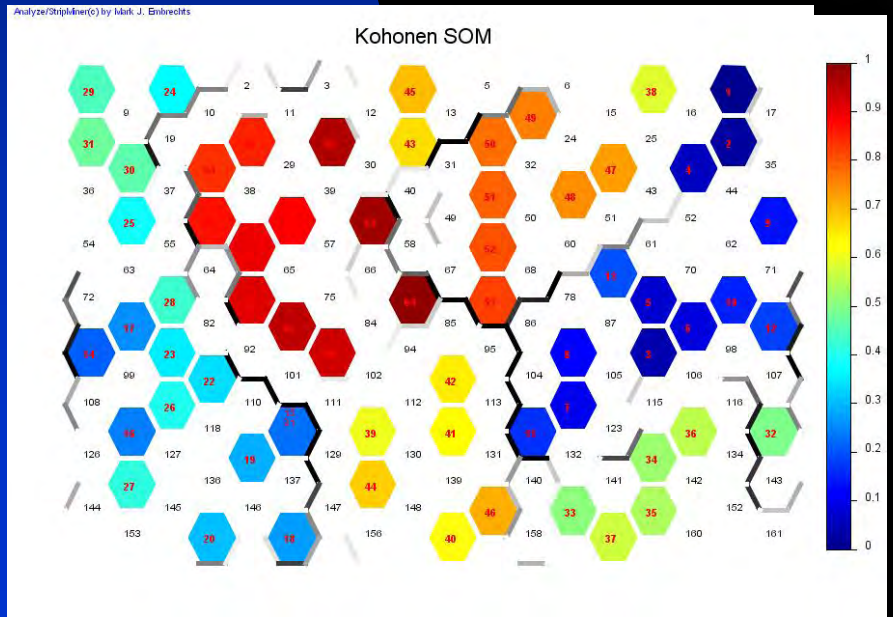
Unsupervised learning - neural network 10 x 10 node projection

Projects high-dimensional input data onto two-dimensional SOM

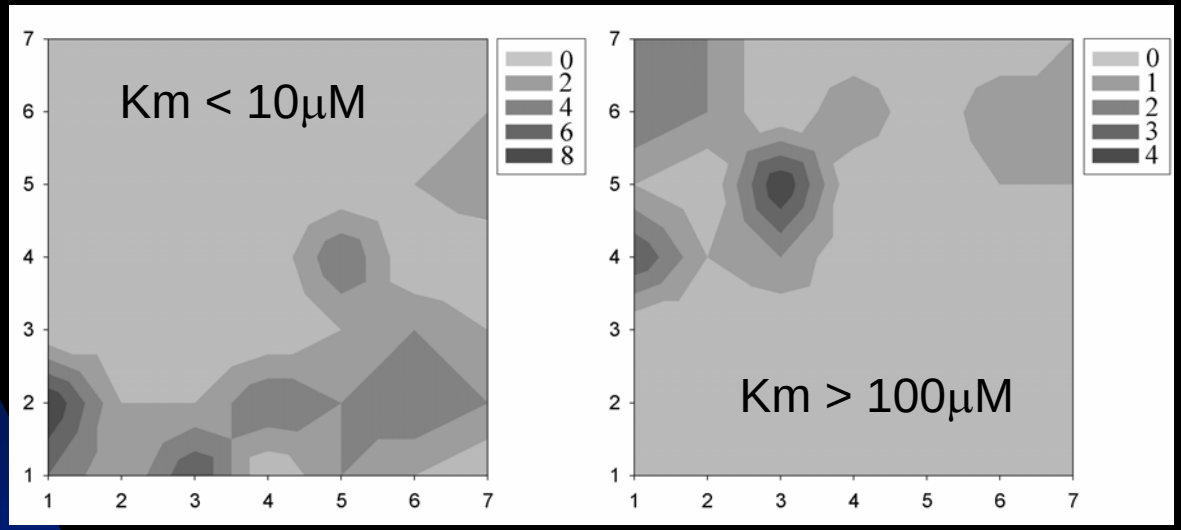
Preserves the topology of the input data

Cluster visualization



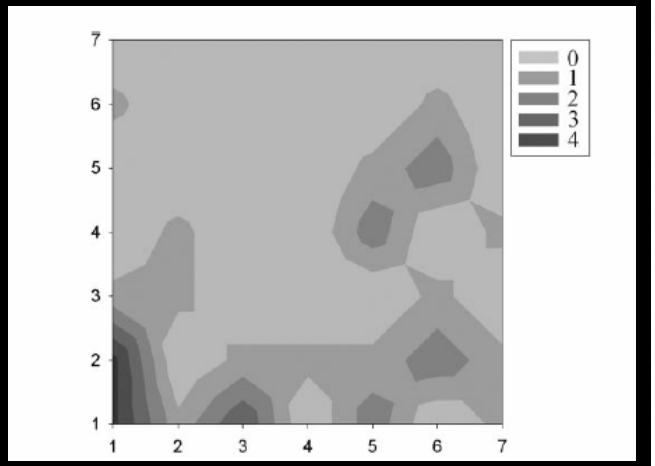


CYP3A4 binding SOM N= 126 molecules

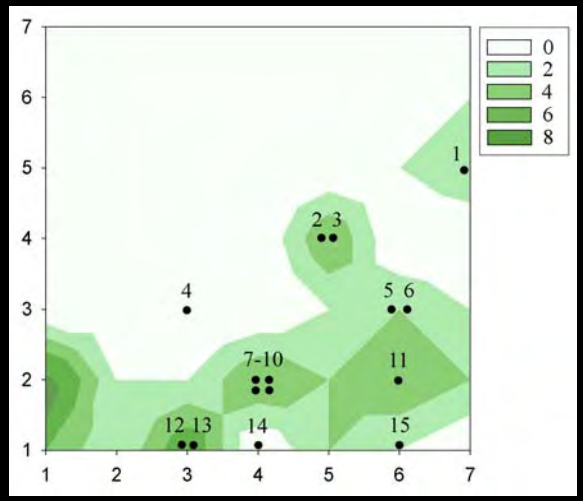


87% - 94% of molecules located correctly in CYP3A4 low Km region of map

Testing with 33 CYP3A4 binders



15 CYP3A4 inhibitors



Balakin et al DMD 32: 1183-1189 (2004)

CYP3A4 Metabolic Intermediate Complex (MIC) formation

Mechanism-based inhibitor = binds to the active site, then becomes catalytically activated by the enzyme

Activated form of the molecule irreversibly binds to the enzyme to remove it from the active enzyme pool.

Some mechanism-based inhibitors cause irreversible inhibition by forming a MIC with the heme

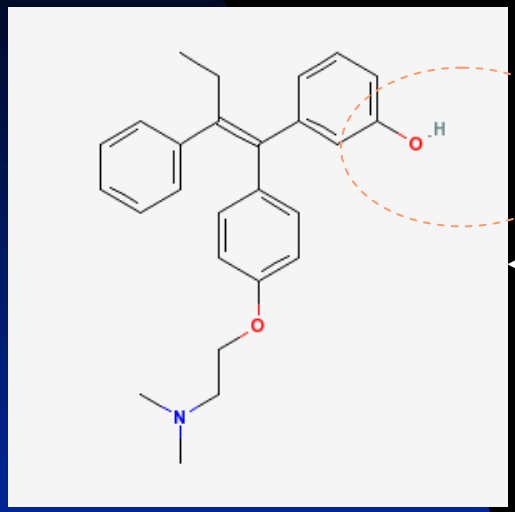
Inactive CYP could lead to misinterpretation of DDI data

Primary, secondary or tertiary amines, or methylenedioxy constituents are prerequisites for MIC compounds (Franklin, 1977).

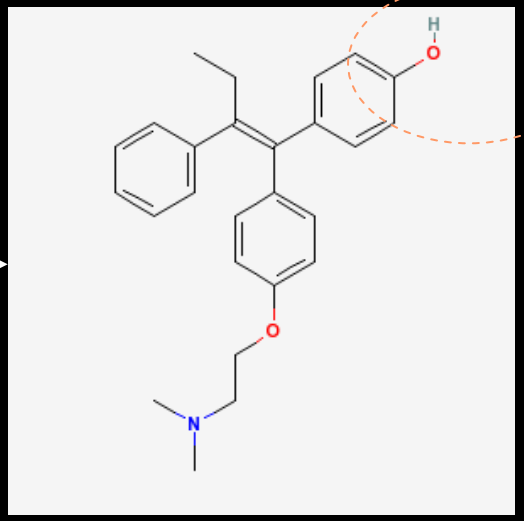
No previous attempts to computationally model MIC formation

Subtle differences - impact on MIC Formation

3-hydroxytamoxifen

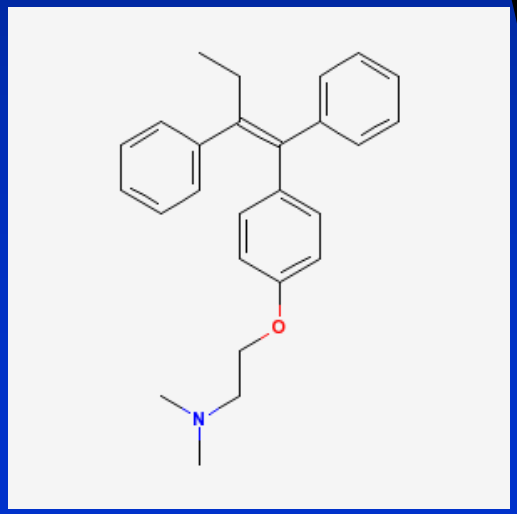


4-hydroxytamoxifen



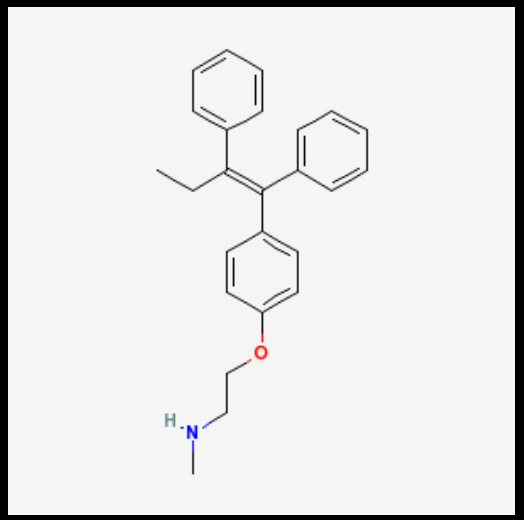
Do not form MIC

Tamoxifen



Form MIC

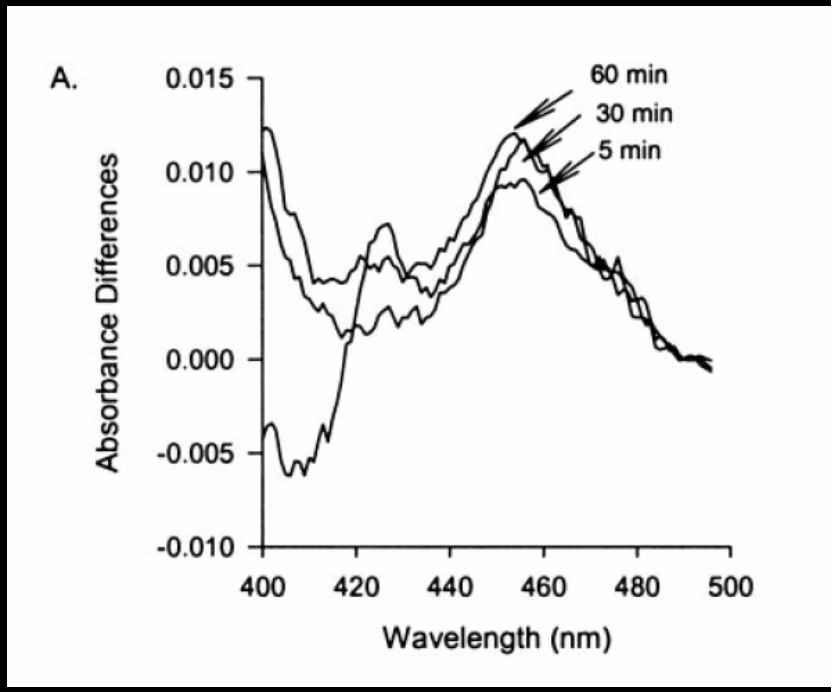
N-desmethyltamoxifen



Molecules tested

54 molecules assessed for MIC formation with recombinant CYP3A4 (+b_5) in vitro (27 MIC +, 27 MIC -)

- Antibiotics
- Calcium channel blockers
- CNS drugs
- HIV protease inhibitors
- Anticancer
- Miscellaneous



Used dual wavelength spectroscopy scanning from 380-500nm
Difference spectra calculated at 490nM vs 452nM at a specific time
Extinction coefficient 65mM^{-1}

Jones et al 2007, DMD in press

Simple property analysis

Generated calculated LogP and molecular weight with ChemDraw for excel

t-test with SPSS

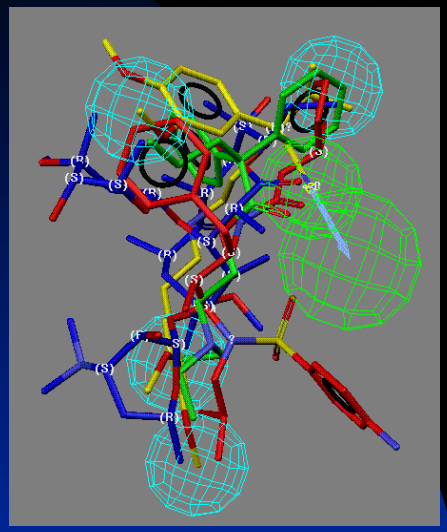
	MIC Mean	MIC SD	MIC range	**Non-MIC Mean**	Non-MIC SD	Non-MIC range
MWT	**472 ***	174	263.4-798	**308**	137	133.2-670.9
cLogP	**3.92**	1.44	1.44-6.81	**3.86**	1.53	1.24-7.05

* $p < 0.05$

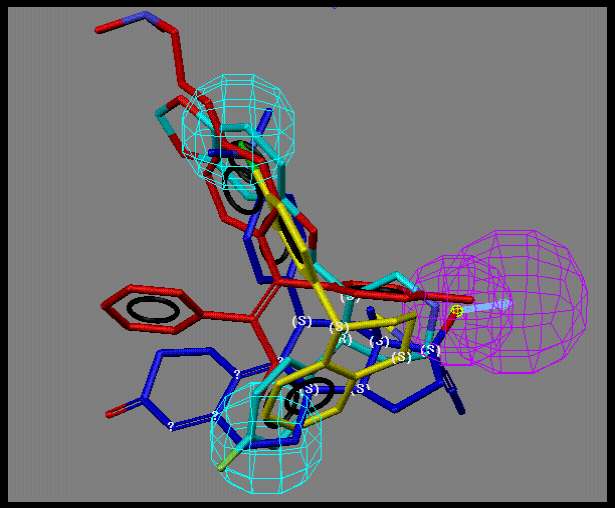
Across all 54 compounds:
number of rotatable bonds & molecular weight ($r^2 = 0.68$)
number of hydrogen bond acceptors & molecular weight ($r^2 = 0.75$)
number of hydrogen bond donors & molecular weight ($r^2 = 0.42$)

Jones et al 2007, DMD in press

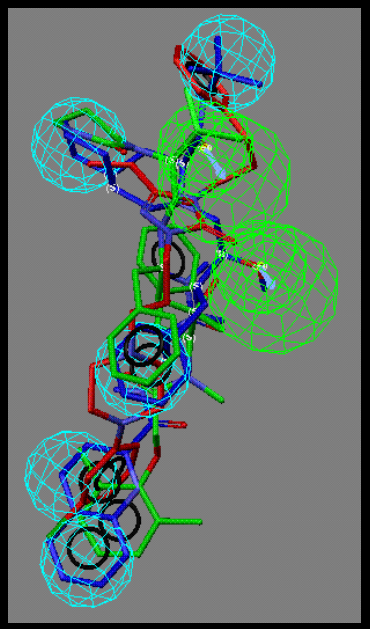
Initial qualitative models: CYP3A4 Metabolite Intermediate Complex



MIC forming compounds



non-MIC forming compounds



non-MIC forming compounds which inactivate CYP3A4

Generated with Catalyst HIPHOP (Accelrys)
Blue spheres = hydrophobic, green feature = hydrogen bond acceptor, Purple spheres = hydrogen bond donor.

Jones et al 2007, DMD in press

Metabolite Intermediate Complex - Recursive partitioning

ChemTree (GoldenHelix) –single tree and 100 random trees (cutoff 0.5) using ChemTree path length descriptors alone

Single tree 87% correct, 100 tree 91% correct with Chemtree descriptors 100 tree model with Cerius2 and ChemTree descriptors

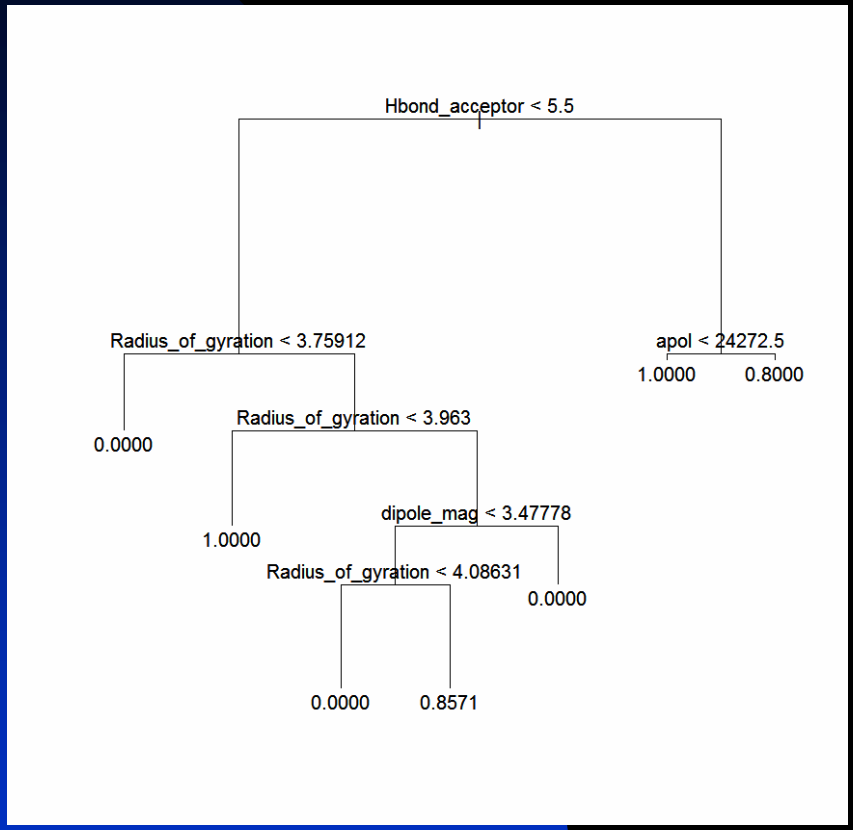
Cerius^2 CSAR (Accelrys) – internally validated 10 fold (74%), 5 fold (80%) and 2 fold (76%) cross validation

Tree function in R – five fold cross validation 96% correct

All models tested with 12 compounds from the literature

Jones et al 2007, DMD in press

R tree model for CYP3A4 (+b_5) MIC formation



model had 96 % prediction accuracy for the 54 compounds in the training set.

Logistic Regression

$$\Pr(MIC = +) = \frac{e^{-2.47+0.59\times \text{Hbond\_Acceptor}}}{1+e^{-2.47+0.59\times \text{Hbond\_Acceptor}}}$$

The hydrogen bond acceptor descriptor was the most important predictor (p-value = 0.002).

The logistic regression model had 80 % prediction accuracy for the 54 compounds in the training set.

Jones et al 2007, DMD in press

Summary of descriptors & training and test set correlations

Prediction Model	Selected Descriptors	Training (n=54)	Test set (n=12)
Pharmacophores	Hydrophobic feature and Hydrogen bond acceptor	NA	NA
ChemTree (1 and 100 tree models)	Hydrogen bond acceptor and hydrophobic	49/54 = **91%**	10/12 = **83.3%**
ChemTree (100 tree) and Cerius^2 descriptors	Hydrogen bond acceptor, hydrophobic and Radius of Gyration,	49/54 = **91%**	10/12 = **83.3%**
Cerius^2 CSAR Tree	Hydrogen bond acceptor, Hydrogen bond donor, area, and Dipole magnitude	54/54 = **100%**	10/12 = **83.3%**
Tree in R	Hydrogen bond acceptor, radius of gyration, sum of atomic polarization, and dipole magnitude	52/54 = **96%**	11/12 = **91.6%**
Linear model in R	Hydrogen bond acceptor,	43/54 = **80%**	11/12 = **91.6%**

Jones et al 2007, DMD in press

Collected > 80 general rules for phase I and II metabolism

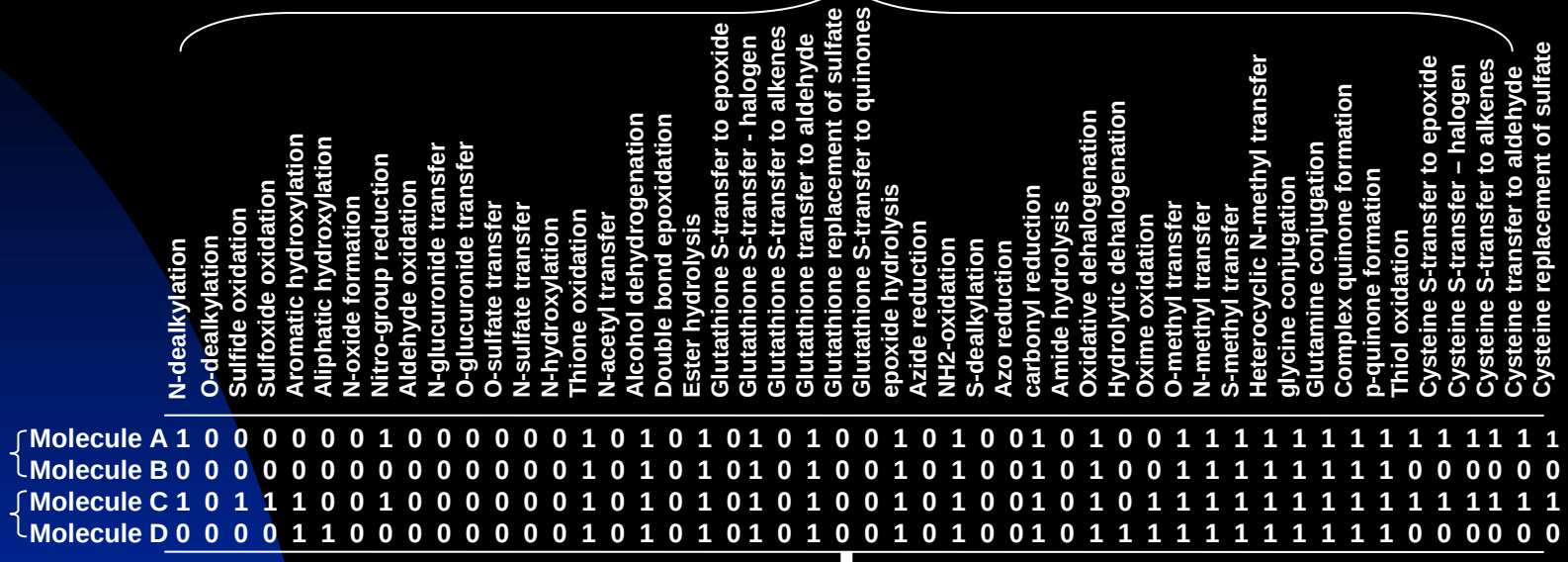
N-dealkylation
O-dealkylation
S-dealkylation
sulfide oxidation
sulfoxide oxidation
aromatic hydroxylation
aliphatic hydroxylation
N-oxide formation
Nitro-group reduction
Double bond peroxidation
Hydroxyl-carbonyl oxidation
aldehyde oxidation
Double bond formation (desaturation)
N-hydroxylation
Thione oxidation
N-acetyl transfer
oxidative deboronation
Double bond epoxidation
ester hydrolysis
epoxide hydrolysis
Azide reduction
oxidative deamination
Glutathione S-transfer to benzyl

Azo reduction
carbonyl reduction
Amide hydrolysis
Oxidative dehalogenation
Hydrolytic dehalogenation
Oxime oxidation
Complex quinone formation
o-quinone formation
p-quinone formation
Thiol oxidation
Phosphate hydrolysis
Phosphite hydrolysis
Phosphorothioate to phosphate
Phosphite oxidation
Sulfoxide reduction
Carboxyl reduction
Carbonyl halide hydrolysis
Decarboxylation
peptide hydrolysis
unsaturated bond hydratation
transamination
N-formyl transfer
O-phosphate transfer
O-acetyl transfer

N-glucuronoside transfer
O-glucuronoside transfer
O-sulfate transfer
N-sulfate transfer
S-glutathione transfer
Glutathione S-transfer to epoxide
Glutathione S-transfer - halogen
Glutathione S-transfer to alkenes
Glutathione transfer to aldehyde
Glutathione replacement of sulfate
Glutathione S-transfer to quinones
O-methyl transfer
N-methyl transfer
S-methyl transfer
Heterocyclic N-methyl transfer
glycine conjugation
Glutamine conjugation
Cysteine S-transfer to epoxide
Cysteine S-transfer - halogen
Cysteine S-transfer to alkenes
Cysteine transfer to aldehyde
Cysteine replacement of sulfate
Glutathione S-transfer to nitroarenes
Cysteine S-transfer to benzyl

Metabolite prioritization

Known molecules queried with n biotransformation rules to annotate observed metabolites as a binary bit string



Add path length molecular descriptors

Generate machine learning algorithm to predict the metabolite binary bit string for each input molecule

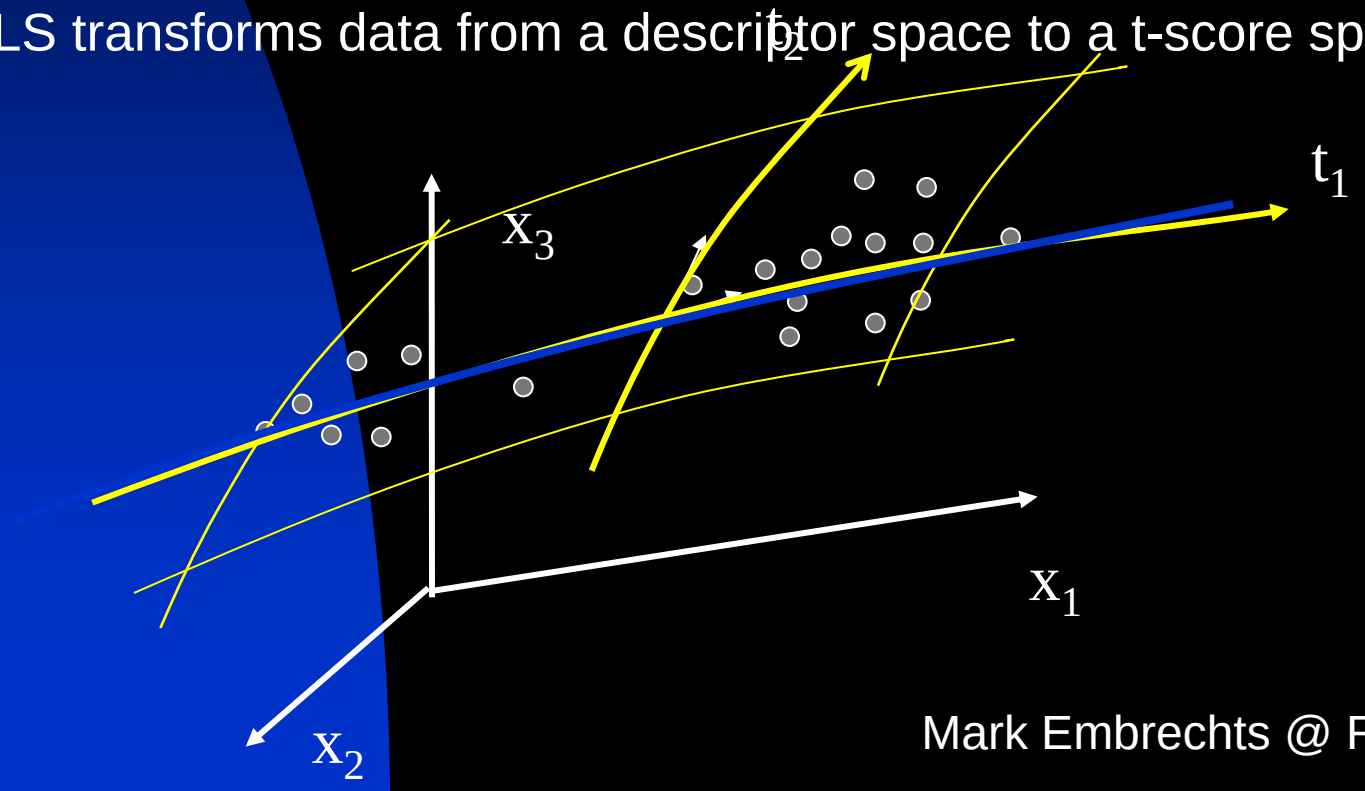
Collected over 300 molecules with human drug metabolism data
Used as a binary training set

Ekins, in Computer applications in pharmaceutical research and development, Wiley 2006

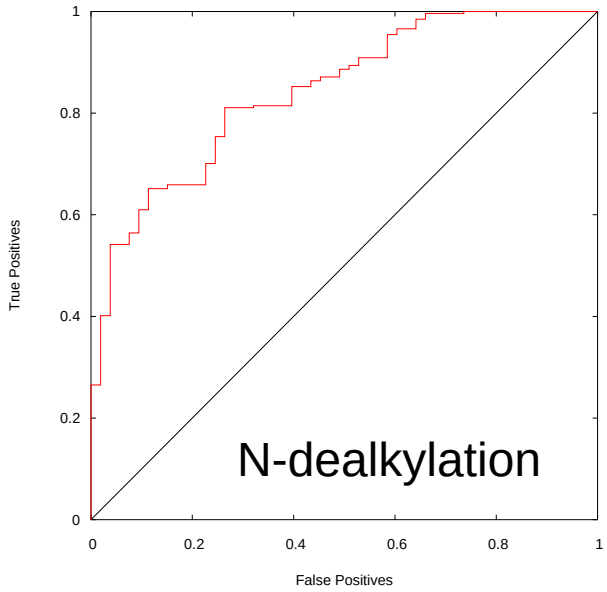
Embrechts and Ekins, DMD 35: 325-327, 2007

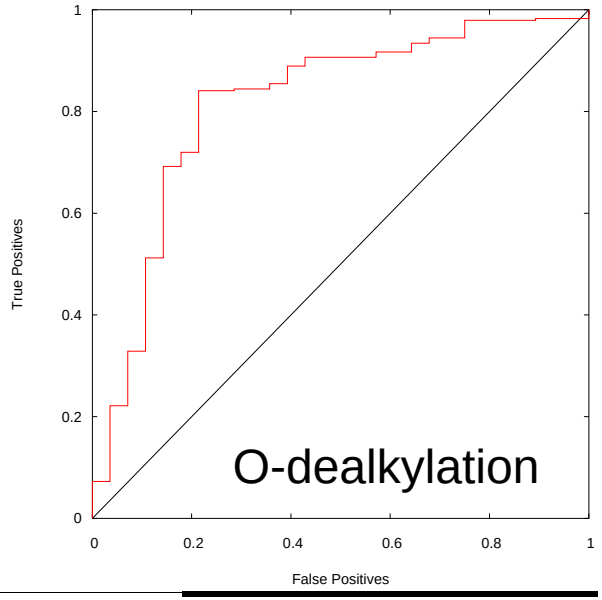
Machine learning -Kernel PLS (K-PLS)

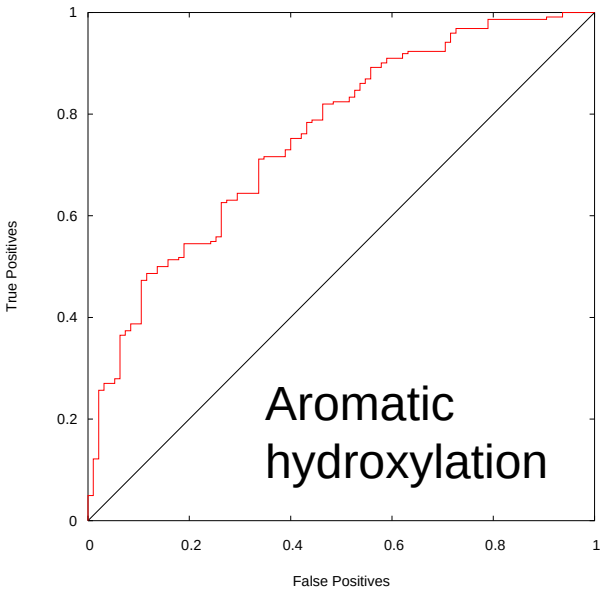
- Direct Kernel PLS is PLS with the kernel transform as a preprocessing step
- Consider K-PLS as a “better” nonlinear PLS
- K-PLS gives almost identical (but more stable) results as support vector machines (SVMs)
  - easy to tune (5 latent variables)
  - unlike SVMs there is no patent on K-PLS
- K-PLS transforms data from a descriptor space to a t-score space

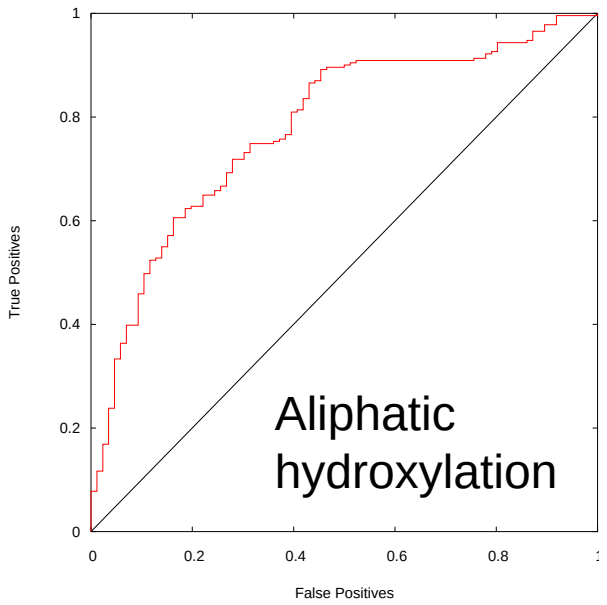


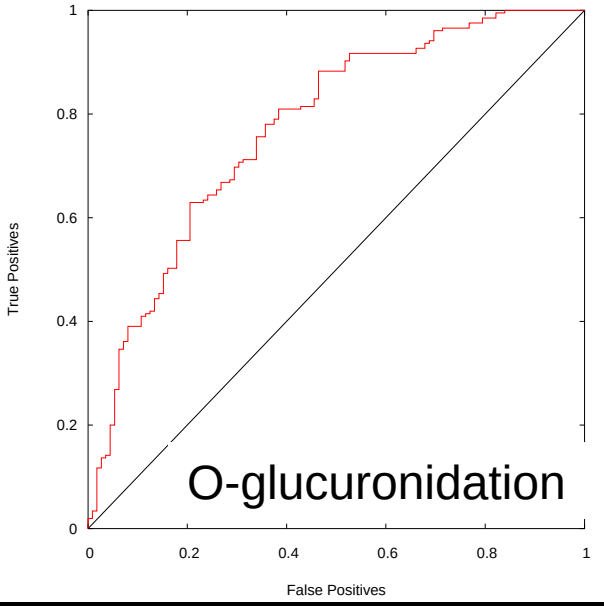
K-PLS results metabolite prediction

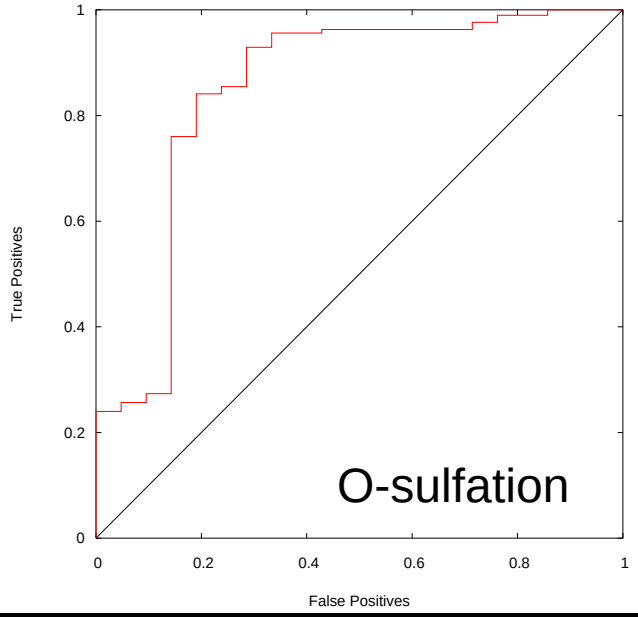






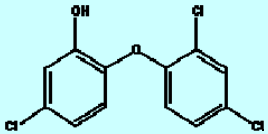






Ekins, in Computer applications in pharmaceutical research and development, Wiley 2006
Embrechts and Ekins, DMD 35: 325-327, 2007

MetaDrug: A hybrid method

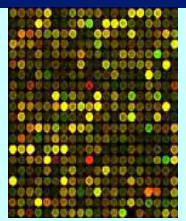


Molecules (sdf, mol)

Metabolite rules

QSAR models

Prioritize metabolites



HT-data

Pathway database/maps

Visualization tools

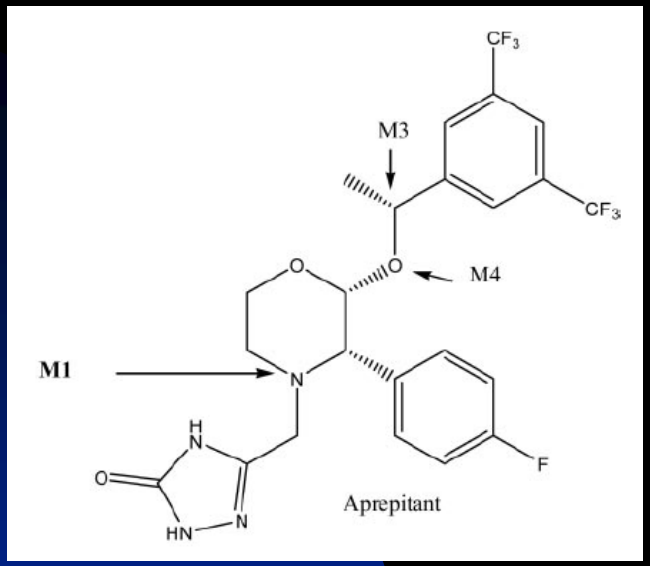
Visualize predicted & empirical data

Select most promising compounds

Software written in Perl
Major components:
CambridgeSoft ChemDraw ActiveX, GoldenHelix ChemTree, Accelrys Accord Oracle
Client – server software

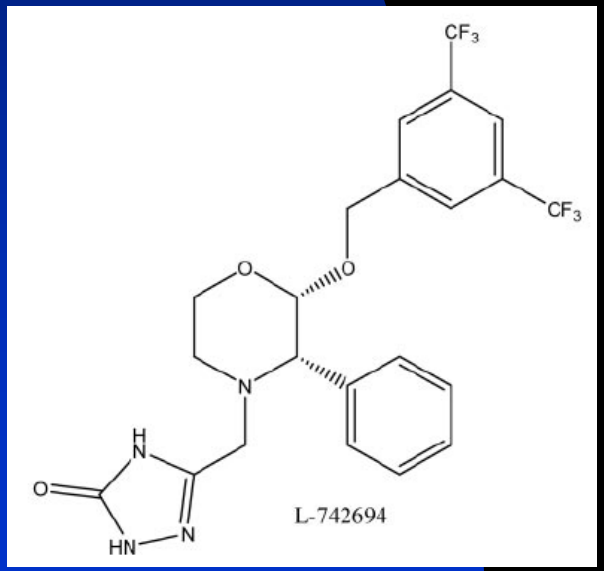
Launched Sept 2004 by GeneGo - Patent Pending

Combined Approach to Metabolism and Toxicity Assessment



Molecule predicted to have a relatively high affinity for:
CYP3A4 Km (predicted, 15 uM; actual ~10 uM, similarity score =0.78)
CYP3A4 Ki (predicted, 13.5 uM; actual 10 uM, similarity 0.78)
PXR (predicted to bind, 0.90 similarity score = 0.77)

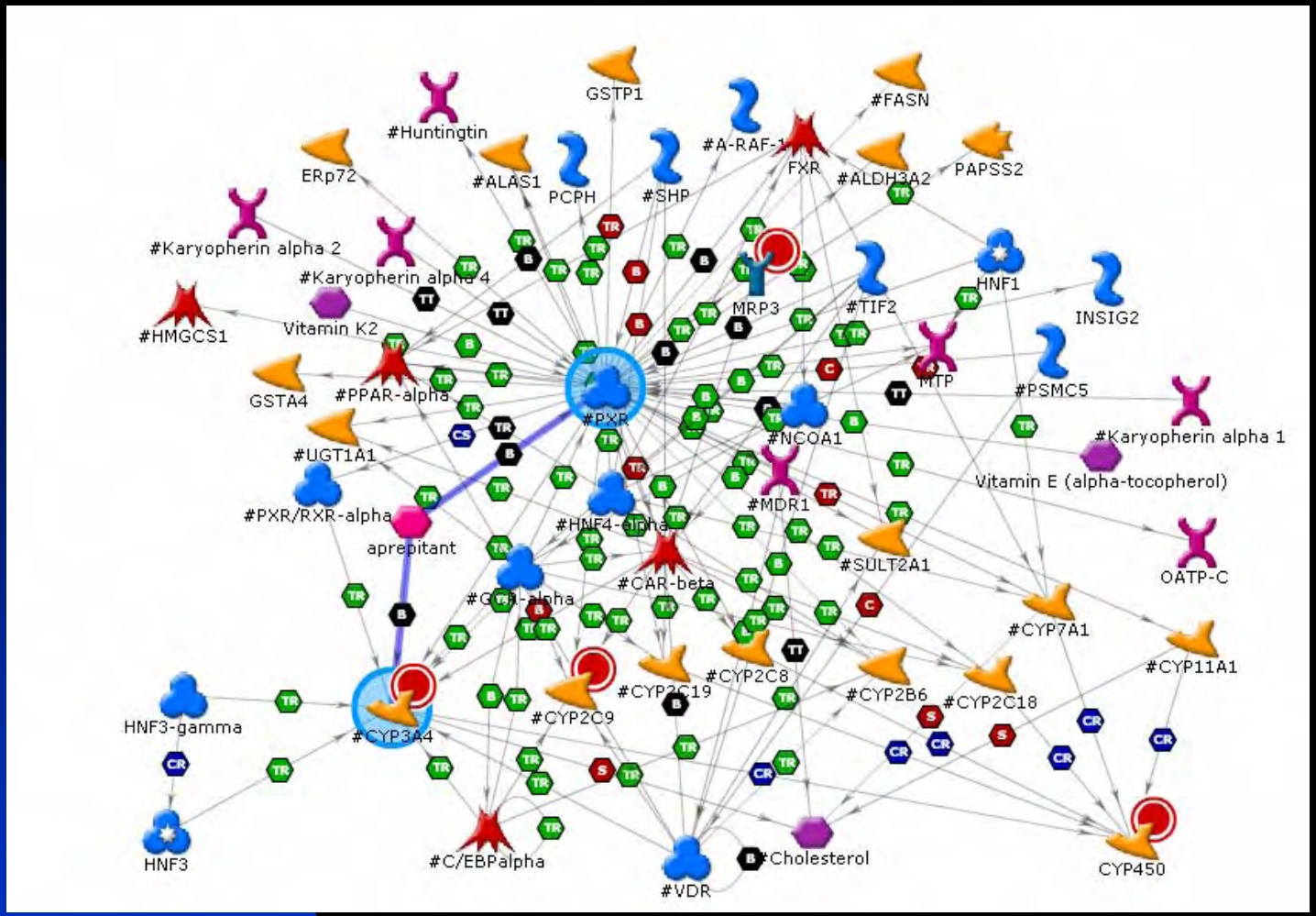
Package insert – Known CYP3A4 & P-gp inducer



Molecule predicted to have a relatively high affinity for:
CYP3A4 Km (predicted, 14.8 uM; similarity score =0.75)
CYP3A4 Ki (predicted, 8.1 uM; similarity 0.78)
PXR (predicted to bind, 0.58 similarity score = 0.77)

Ekins et al., *Drug Metab Dispos*, 34: 495-503 (2006)

Predicted Interactions and Microarray Data



Autoexpand Network

Data from Hartley et al *Mol Pharmacol* 65 (2004) 1159-1171
Rats treated with L-742694 potent PXR agonist – appears to increase expression of metabolizing enzymes and transporters – increasing clearance?

Ekins et al., *Drug Metab Dispos*, 34: 495-503 (2006)

Conclusions

Use computational methods to screen virtual and real compound libraries

Complexity in prediction of multiple molecules binding simultaneously & location

- Molecule interactions, molecule –water-molecule interactions

Need for new approaches

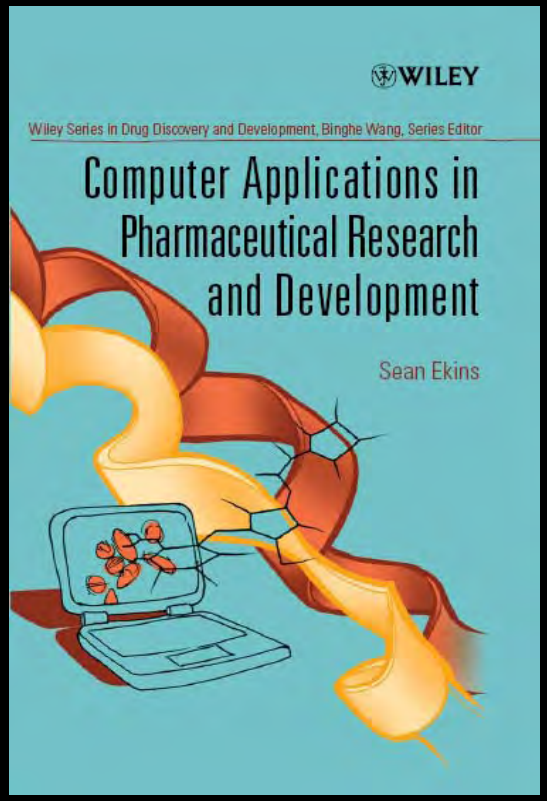
- Statistical models limited by training set
- Understand extrapolations
- Need for more generalizable rules for CYPs
- Approaches that combine regiospecificity, affinity and lability

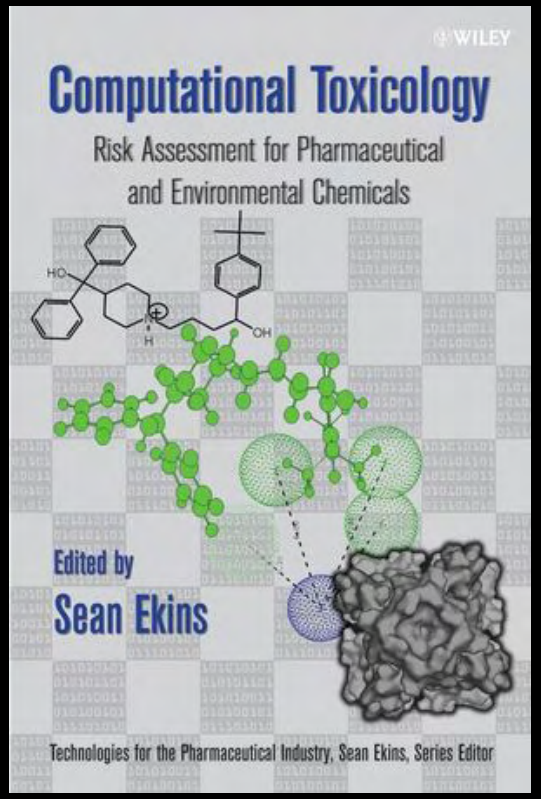
Models for rat and mouse enzymes

Integration of computational models with in vitro methods, model rebuilding

Acknowledgments

- Konstantin Balakin (ChemDiv)
- Mark Embrechts (RPI)
- David Jones (IU)
- Lang Li (IU)
- Steve Hall (IU)
- David Stresser
- Andrew Williams





- Past colleagues at Lilly including Jim Wikel, Steve Wrighton